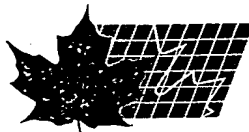


A.D

S

OPIC
OFFICE DE LA PROPRIÉTÉ
INTELLECTUELLE DU CANADA



CIPO
CANADIAN INTELLECTUAL
PROPERTY OFFICE

PCT / CA 99/00637
22 JUL 1999 (22.07.99)

CA 99/00637

REC'D 05 AUG 1999	
WIPO	PCT

*Bureau canadien
des brevets*

*Canadian Patent
Office*

Certification

Certification

La présente atteste que les documents
ci-joints, dont la liste figure ci-dessous,
sont des copies authentiques des docu-
ments déposés au Bureau des brevets.

This is to certify that the documents
attached hereto and identified below are
true copies of the documents on file in
the Patent Office.

Specification as originally filed with Application for Patent Serial No: 2,237,704, on July 14, 1998, by **THE UNIVERSITY OF BRITISH COLUMBIA**, assignee of John Smit, for "Cleavage of *Caulobacter* Produced Recombinant Fusion Proteins".

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Agent certificateur / Certifying Officer
July 22, 1999

Date



Industrie
Canada

Industry
Canada

(CIPO 68)

Canada

ABSTRACT

This invention provides a method of cleaving target proteins from Caulobacter S-layer protein under mild acid
5 conditions whereby a fusion protein secreted by the Caulobacter and comprising the target protein and at least a Caulobacter S-layer secretion signal may be cleaved at a aspartate-proline dipeptide without solubilizing the fusion
10 protein. This method may be carried out while the fusion protein is in an insoluble aggregate which facilitates recovery of the protein. This invention also provides a method of preparing a DNA construct for expression of the fusion protein and a method of preparing the fusion protein.

**CLEAVAGE OF CAULOBACTER PRODUCED
RECOMBINANT FUSION PROTEINS**

FIELD OF INVENTION

5

This invention relates to the expression and secretion of recombinant fusion proteins from Caulobacter wherein a heterologous polypeptide is fused with all or part of the surface layer protein (S-layer protein) of the bacterium.

10

BACKGROUND OF THE INVENTION

Many bacteria assemble layers composed of repetitive, regularly aligned, proteinaceous sub-units on the outer surface of the cell. These layers are essentially two-dimensional paracrystalline arrays, and being the outer molecular layer of the organism, directly interface with the environment. In Caulobacter, the S-layer protein is synthesized by the cell in large quantities and the S-layer completely envelopes the cell and thus appears to be a protective layer.

Caulobacter are natural inhabitants of most soil and freshwater environments and may persist in waste water treatment systems and effluents. The bacteria alternate between a stalked cell that is attached to a surface, and an adhesive motile dispersal cell that searches to find a new surface upon which to stick and convert to a stalked cell. The bacteria attach tenaciously to nearly all surfaces and do so without producing the extracellular enzymes or polysaccharide "slimes" that are characteristic of most other surface attached bacteria. Caulobacters have simple requirements for growth. The organism is ubiquitous in the environment and has been isolated from oligotrophic to mesotrophic situations. They are known for their ability to tolerate low nutrient level stresses, for example, low phosphate levels.

-2-

All of the freshwater Caulobacter that produce an S-layer are similar and have S-layers that are substantially the same under electron microscopy. The layers are hexagonally arranged in all cases, with a similar centre-centre dimension (see: Walker, S.G., et al. (1992). "Isolation and Comparison of the Paracrystalline Surface Layer Proteins of Freshwater Caulobacters" J. Bacteriol. 174: 1783-1792).

16S rRNA sequence analysis of several S-layer producing Caulobacter strains show that they group closely (see: Stahl, D.A. et al. (1992) "The Phylogeny of Marine and Freshwater Caulobacters Reflects Their Habitat" J. Bacteriol. 174: 2193-2198). DNA probing of Southern blots using the S-layer gene from C. crescentus CB15 identifies a single band that is consistent with the presence of a cognate gene (see: MacRae, J.D. and, J. Smit. (1991) "Characterization of Caulobacters Isolated from Wastewater Treatment Systems" Applied and Environmental Microbiology 57:751-758). Furthermore, antisera raised against the S-layer protein of CB15 reacts against the S-layer protein of other Caulobacter (see: Walker, S.G. et al. (1992) [*supra*]). All S-layer proteins isolated from Caulobacter may be substantially purified using the same methods. All strains appear to have a polysaccharide species which may be required for S-layer attachment (see: Walker, S.G. et al. (1992) [*supra*]).

The S-layers elaborated by freshwater isolates of Caulobacter are visibly indistinguishable from the S-layer produced by Caulobacter strains CB2 and CB15. The S-layer proteins from the latter strains have approximately 100,000 m.w. although sizes of S-layer proteins from other species and strains will vary. The hydrophillic S-layer protein has been characterized both structurally and chemically. It is composed of ring-like structures spaced at 22 nm intervals arranged in a hexagonal manner on the outer

-3-

membrane. The S-layer is bound to the bacterial surface and may be removed by low pH treatment or by treatment with a calcium chelator such as EDTA.

5 The similarity of S-layer proteins in different strains of Caulobacter permits the use of a cloned S-layer protein gene of one Caulobacter strain for retrieval of the corresponding gene in other Caulobacter strains (see: Walker, S.G. et al (1992) [*supra*]; and MacRae, J.D. et al
10 (1991) [*supra*]).

Expression of a heterologous polypeptide as a fusion product with the S-layer protein of Caulobacter provides advantages not previously seen in systems for production of
15 recombinant fusion proteins using other organisms such as E. coli and Salmonella. All known Caulobacter strains are believed to be harmless and are nearly ubiquitous in aquatic environments. In contrast, many Salmonella and E. coli strains are pathogens. Consequently, expression and
20 secretion of a heterologous polypeptide using Caulobacter as a vehicle has the advantage that the expression system will be stable in a variety of outdoor environments and may not present problems associated with the use of a pathogenic organism. Furthermore, Caulobacter are natural
25 biofilm forming species and may be adapted for use in fixed biofilm bioreactors. The quantity of S-layer protein that is synthesized and is secreted by Caulobacter is high, reaching 12% of the cell protein.

30 There is an existing need to produce pure proteins and peptides in an economical manner and in a manner that minimizes or simplifies the purification steps needed after fermentation. Key commercial areas include the production of recombinant human and animal therapeutic antibiotic and
35 vaccine peptides, industrial enzymes, protein polymers, and antibacterial enzymes for foodstuffs. Many of these commercial applications require low production costs and

-4-

there are few expression systems available that can meet such cost restraints. In addition, there are numerous research applications where rapid methods to produce and purify proteins are needed to facilitate the discovery stage. This is especially true where there is a desire to express a large number of proteins with unknown function (from a collections of cloned cDNA's, for example) or a large number of variants of a single protein, (for example, resulting from site directed mutagenesis) in a search for variants with improved properties.

Generally, proteins must be secreted to be produced at low cost. The primary reason is the much reduced cost of purification of the target protein from cell material. However, even for secreted proteins, simple methods of separating the product from spent culture and cells are important for cost reduction and ease of use.

International patent application published as WO 97/34000 on September 18, 1997 describes the expression and secretion of recombinant proteins from Caulobacter in which the recombinant protein is a fusion of all or part of Caulobacter S-layer protein with a heterologous protein of interest (also see: Bingle, W.H., et al 1997¹ "Linker Mutagenesis of the Caulobacter us S-layer protein: Toward a Definition of an N-terminal Anchoring Region and a C-terminal Secretion Signal and the Potential for Heterologous Protein Secretion". J. Bacteriol. 179:601-611).

The Caulobacter S-layer secretion apparatus is in the category of "Type 1" secretion usually found in pathogenic bacteria and noted for its ability to secrete a wide variety of proteins including large and hydrophilic proteins. The Caulobacter protein secretion system is particularly useful to secrete recombinant proteins.

-5-

The Caulobacter S-layer Type 1 secretion pathway requires only a C-terminal secretion signal, typically comprising about 200 amino acids at the end of the protein. The export mechanism is capable of tolerating a wide variety of foreign proteins. Recombinant proteins may be conveniently produced as fusion proteins with the target protein being fused to the C-terminal secretion signal. Depending on the application, it may be desirable to remove the secretion signal following secretion. Not removing the secretion signal may be an approach suitable for many subunit vaccine applications, where the remaining S-layer protein serves as a carrier.

A unique and desirable feature of fusion proteins produced by the Caulobacter S-layer protein secretion system is that they form insoluble aggregates in the culture medium. This is apparently a consequence of the S-layer sequences associated with secretion signal and reflects the fact that the protein normally self-assembles into a two dimensional crystalline layer on the bacterium's surface. These aggregates are visible to the naked eye and are readily collected by simple filtration. With simple water wash steps, residual bacterial cells are readily flushed away. It is routinely possible to achieve a protein purity of 90% or better with this simple purification procedure.

DESCRIPTION OF THE PRIOR ART

Most current protein purification systems for recombinant proteins produced by bacteria rely upon an affinity matrix to achieve separation of the target protein and to concentrate the protein for subsequent steps of purification. To accomplish this, genes for recombinant proteins are commonly constructed so that they contain affinity tags, which are protein sequences that will bind to an affinity matrix. Some commonly used systems are:

-6-

- 1) Glutathione S-transferase (GST) tag, which binds to glutathione-sepharose matrices.
- 2) Maltose binding protein (MBP) tag, which binds to amylose matrices.
- 3) Multiple tandem histidine residues (e.g. "His-6") tag, which binds to Nickel-derivatized solid matrices.
- 4) Protein A tag, which binds to Immunoglobulin IgG-derivatized sepharose or comparable matrices.

Prior art techniques were typically developed so that removal of a target protein does not disrupt the tag and matrix association. Instead, enzymes that cleave specific sequences of amino acids are employed. The enzyme cleavage sequence is positioned between the tag and the desired recombinant protein and enzymatic cleavage is effected directly on the matrix with attached fusion protein. If a secretion signal is used, the cleavage site is usually positioned such that the secretion signal is separated from the target recombinant protein during the cleavage step. The matrix is regenerated for re-use only after the target recombinant protein has been purified away from the matrix. Typical enzymes used in these methods are Factor Xa, enterokinase and collagenase.

Chemical cleavage is generally not used because the conditions required for cleavage will disrupt the binding of affinity tag and matrix or destroy the matrix. When chemical cleavage is used with recombinant fusion proteins to cleave target protein from a secretion signal and/or affinity tag, solubilization and denaturation processes are generally employed. The expectation is that complete or nearly complete unfolding of the protein is a prerequisite for effective cleavage.

-7-

Mild-acid cleavage is predicated on the inclusion, by happenstance or design, of the acid-sensitive aspartate-proline dipeptide at a desired site for cleavage. The recombinant fusion protein is exposed to conditions
5 that solubilize and/or completely denature the protein prior to cleavage. The chaotropic agent guanidine hydrochloride (used at 6-7 M) is commonly employed to denature and solubilize the protein prior to, or at the same time as acid treatment. Alternately, high
10 concentrations of acids that also serve as solubilizing agents (as examples: 70-90% formic acid, acetic acid [10%] pyridine, or relatively high concentrations of HCL (60 mM or more) are employed. Because such conditions would disrupt a tag/affinity matrix association, direct cleavage
15 of an affinity tag from the target protein while a protein remains associated with an affinity matrix is not attempted.

SUMMARY OF INVENTION

20 This invention is based on the unexpected discovery that recombinant fusion proteins produced by the Caulobacter S-layer protein secretion system can be cleaved under mild-acid conditions without solubilization of the fusion
25 protein being required. This cleavage may be accomplished when the fusion protein is present as the insoluble aggregate typically formed by Caulobacter S-layer protein. Cleavage occurs at aspartate-protein dipeptides which may be in the heterologous protein portion of the fusion
30 protein or native to the Caulobacter S-layer portion. The dipeptide may also be placed at a desired location for cleavage by engineering DNA encoding the fusion protein to express the dipeptide at the desired location. Typically, the desired location for cleavage will be at or near the
35 junction of the heterologous (target) protein and the Caulobacter S-layer portion comprising the Caulobacter secretion signal such that a cleavage product will be the

-8-

target protein in its entirety and preferably free of extraneous amino acids.

5 The current invention makes it possible to cleave a heterologous (target) protein from the S-layer protein portion using only mild-acid conditions, even while the fusion protein is in an aggregated form. These cleavage conditions do not result in significant solubilization of the S-layer protein portion. In essence, the S-layer
10 aggregation phenomenon functions like an affinity matrix.

This invention provides a method of cleaving a fusion protein consisting of a first component comprising all or part of a Caulobacter S-layer protein including a
15 Caulobacter C-terminal secretion signal, and a second component comprising a heterologous polypeptide expressed and secreted from Caulobacter wherein the fusion protein comprises at least one aspartate-proline dipeptide, and wherein the method comprises the step of combining said
20 fusion protein with an acid solution of a strength insufficient to solubilize the fusion protein for a time sufficient for cleavage of the fusion protein at said aspartate-proline. The acid solution may have a pH of from about 1.5 to about 2.5, preferably about 1.65 - 2.35.
25 Preferred pH conditions may be achieved using an acid equivalent in the range of about 5 to about 20 mM HCL. The method is typically carried out at a temperature in the range of approximately room temperature to about 50°C.

30 This invention also provides a method of preparing a DNA construct suitable for expression of a fusion protein suitable for use in the method of this invention, comprising joining an upstream DNA segment comprising DNA for a heterologous protein of interest and a downstream DNA
35 segment for a Caulobacter C-terminal secretion signal which does not encode an aspartate-proline dipeptide, wherein the upstream segment comprises a sequence encoding an

-9-

aspartate-proline dipeptide at or near a junction between said upstream and downstream segments.

5 This invention also provides a method of preparing a fusion protein, comprising the steps of expressing a DNA construct as described above in Caulobacter and recovering said fusion protein once secreted by the Caulobacter.

10 Once cleavage is accomplished according to this invention, the S-layer portion comprising the Caulobacter secretion signal may remain as an insoluble aggregate. If the target protein is soluble, the S-layer portion may be easily separated from the target recombinant protein by simple centrifugation or filtration methods. Thus the system
15 functions in a manner analogous to a Tag/affinity matrix system except that here, the affinity "tag" is the means of producing the insoluble matrix. In addition, this "matrix" is resistant to the effects of the acid treatment, allowing direct cleavage of the target recombinant protein. In this
20 way, a very inexpensive chemical cleavage method can be employed to economically retrieve recombinant proteins from a bacterial fusion protein. In contrast to the cost of most affinity matrices, there is little expense associated with the use of the S-layer secretion signal as the matrix
25 as it is simply a part of the fermentation/secretion process.

DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

30 Production of Recombinant Fusion Proteins Using the Caulobacter S-layer Secretion System

Proteins may be produced using the Caulobacter S-layer Type 1 secretion pathway which requires only the C-terminal
35 secretion signal of the Caulobacter. This signal is the C-terminal portion of the S-layer protein, which typically comprises about 200 amino acids. (See: Bingle, et al (1997))

-10-

[supra]; and, WO 97/34000). Additional Caulobacter S-layer gene sequence upstream from the secretion signal may also be present and is desirable to contribute to aggregate formation of the secreted protein. The additional

5 Caulobacter sequence may constitute most or all of the remainder of the S-layer protein. Typically, the aggregate forms as loose, gel-like clumps of pure protein that can be readily retrieved and separated from the bacteria by simple filtration.

10 Standard techniques (such as methods described in WO 97/34000) may be used to identify the amount of the C-terminal portion of a particular Caulobacter S-layer protein which functions as the secretion signal.

15 The creation of fusion proteins is commonly done by preparing the target gene DNA and fusing it in-frame with the C-terminal region of the S-layer gene. There are numerous possible methods; some examples are:

20 1. **Oligonucleotide Chemical Synthesis.** This involves the design of complementary single strands, complete with desirable restriction endonuclease cut sites at the ends, chemical synthesis of the strands followed by annealing,

25 cloning into a plasmid vector, juxtaposed to an appropriate portion of the C-terminal region of the S-layer gene.

30 2. **Production of the Target Gene DNA by Polymerase Chain Reaction (PCR) Amplification of a Target Sequence.** In this case, appropriate in-frame restriction sites are incorporated into the short oligonucleotides used for amplification of a target sequence, such that the final PCR product can be treated with the appropriate restriction enzymes (to create the restriction site "sticky ends"),

35 followed by cloning into a plasmid vector, juxtaposed to an appropriate portion of the C-terminal region of the S-layer gene.

-11-

3. Adapting Restriction Endonuclease Cleavage Sites that are Native to a Target Protein Gene Sequence for Fusion to the DNA Coding for the C-terminal S-layer Secretion Signal to Accomplish In-frame Expression of a Chimeric Protein.

5 This can be accomplished by direct ligation (although it is uncommon that an appropriate match will occur), or the use of adapter sequences or methods involving blunting of a restriction site and subsequent blunt-end ligation to change expression reading frame or join unlike restriction
10 site sticky ends.

There will be numerous convenient sites for fusion with the C-terminal regions of the S-layer that lead to the successful expression, secretion and aggregation of a
15 recombinant fusion protein. Some example positions are at or near the DNA sites corresponding to amino acids 622, 690, 784, 892 and 907 of the C. crescentus S-layer gene (see: Appendix 1 and, WO 97/34000). Other sites of fusion with the S-layer gene may also be employed. Most often a
20 plasmid vector is designed such that the C-terminal gene segment is resident on a plasmid with appropriate restriction sites placed at the N-terminal junction of the S-layer fragment. Target recombinant protein gene segments are then cloned into those restriction sites. It is
25 typical to prepare initial plasmid constructs that are replicated in E.coli. After a construct is produced, it is typically transferred to a broad host range plasmid which can then be introduced into the appropriate Caulobacter strain by electroporation. Suitable broad host range
30 plasmids can be constructed from (but are not limited to) the IncQ, IncW and IncP1 plasmid incompatibility groups.

The introduction of the aspartate-proline (Asp-Pro) dipeptide at the appropriate site in the fusion protein can
35 be done in several ways. Some examples are:

-12-

1. Incorporating a DNA sequence necessary to express the Asp-Pro dipeptide into the oligonucleotides used to prepare the target sequence, either by oligonucleotide synthesis or PCR methods.

5 2. Preparing a DNA segment with appropriate restriction sites at the termini so that an Asp-Pro dipeptide can be introduced (most often at the junction between S-layer and target gene) after a fusion recombinant S-layer gene has
10 been made.

3. Use of a native Asp-Pro dipeptide in either the target DNA or the S-layer segment. For example, an Asp-Pro dipeptide is located at amino acids 692 and 693 of the C. crescentus S-layer gene (see: Appendix 1 and, WO 97/34000)
15 and is suitable for fusions made at the amino acid site.

The methods described above are not the only methods of creating and expressing the fusion recombinant S-layer
20 proteins, nor is it essential to have the engineered genes resident on a plasmid. For example, the expressed gene may be introduced into the chromosome (using well-known gene insertion or replacement techniques) and still achieve secretion of the recombinant proteins (see WO 97/34000).
25 In some cases it may be desirable to produce recombinant fusion proteins as insertions of heterologous DNA in the middle of the S-layer gene. In such a case, Asp-Pro dipeptide sequences would be engineered at the N and C-termini of the target peptide.

30 All possible codon combinations for Asp-Pro will work but the CCA codon for proline is not preferred due to a likely low amount of the corresponding tRNA in Caulobacter. The following is an approximate usage table for C. crescentus.

35

-13-

TABLE 1

Caulobacter crescentus Codon Usage Table
 [Amino Acid] [Triplet Code] [Frequency Per Thousand]

10	Phe UUU 2.5	Ser UCU 1.2	Try UAU 6.6	Cys UGU 0.6
	Phe UUC 27.0	Ser UCC 8.5	Try UAC 9.6	Cys UGC 5.5
	Leu UUA 0.0	Ser CA 1.2	STOP UAA 0.8	Cys UGA 1.6
	Leu UUG 4.4	Ser UCG 25.7	STOP UAG 0.6	STOP UGG 7.2
	Leu CUU 4.4	Pro CCU 2.5	His CAU 3.2	Arg CGU 7.6
	Leu CUC 15.7	Pro CCC 15.5	His CAC 12.2	Arg CGC 44.7
	Leu CUA 1.1	Pro CCA 0.9	Gln CAA 3.7	Arg CGA 3.0
	Leu CUG 72.3	Pro CCG 27.1	Gln CAG 30.2	Arg CGG 12.1
15	IleAUU 2.4	Thr ACU 1.2	Asn AAU 4.1	Ser AGU 0.8
	Ile AUC 49.0	Thr ACC 37.3	Asn AAC 23.8	Ser AGC 14.9
	Ile AUA 0.3	Thr ACA 0.8	Lys AAA 2.7	Arg AGA 0.4
	Met AUG 25.7	Thr ACG 16.8	Lys AAG 37.9	Arg AGG 1.1
	Val GUU 5.4	Ala GCU 9.5	Asp GAU 11.1	Gly GGU 9.5
	Val GUC 42.7	Ala GCC 84.1	Asp GAC 48.5	Gly GGC 64.8
	Val GUA 1.0	Ala GCA 2.2	Glu GAA 20.5	Gly GGA 2.3
20	Val GUG 30.7	Ala GCG 36.7	Glu GAG 45.4	Gly GGG 7.7

Large quantities (eg. 12% of total cell protein/3% of input organic carbon) of a wide range of proteins can be produced, with yields in the order of 250 mg/liter of batch culture. Fusion proteins with 35 kDa of target peptide are secreted with little difficulty, although proteins with multiple cysteines may be more difficult to express. Post-expression glycosylation of proteins does not occur, an advantage for most peptide expression applications.

-14-

Host Expression Strains

For secretion of recombinant fusion S-layer proteins, the Caulobacter strain will preferably have lost the ability to produce a native S-layer protein, while retaining a fully functional S-layer protein secretion apparatus. This may be done by screening for mutants that have spontaneously become S-layer protein negative; or, by directed genetic manipulation, such as (but not limited to) the insertion of a drug resistance cassette in the middle of the S-layer gene or the substitution of a version of the S-layer gene which has had a sizable internal region deleted from the gene (see: Bingle et al 1997¹ [*supra*]; Bingle et al 1997² "Cell Surface Display of a Pseudomonas aeruginosa PAK Pilin Peptide with the Paracrystalline Layer of Caulobacter crescentus" Molec. Microbiol. 26:277-288; and, Edwards and Smit (1991) "A Transducing Bacteriophage for Caulobacter us Uses the Paracrystalline Surface Layer Protein as a Receptor" J. Bacteriol. 173: 5568-5572). In the case of a genetic manipulation, a common method for producing such strains is to do appropriate modification of a copy of the S-layer gene while on a plasmid and then to use well known gene replacement methods to substitute the modified gene for the native gene in the Caulobacter chromosome (see: Edwards and Smit (1991) [*supra*]).

In the rare case that an entire S-layer gene is used for production of a recombinant protein (via insertion of a target sequence), strains defective in the production of the lipopolysaccharide (LPS) used for S-layer attachment to the bacterial surface can be used. These can be prepared by forcing Caulobacter to grow without exogenous calcium. Under these conditions mutants arise that are uniformly defective in producing a proficient version of the S-layer LPS (see: Walker, S.G. et al (1994) "Characteristics of Mutants of Caulobacter crescentus Defective in Surface

-15-

Attachment of the Paracrystalline Layer" J. Bacteriol. 176: 6312-6323).

5 All Caulobacter S-layer producing strains are suitable for
this technology. One may either isolate the S-layer gene
from a strain (using homology between Caulobacter S-layers
to design probes to detect and clone the S-layer genes) and
adapt the C-terminal region for recombinant protein
expression in a manner similar to that done for C.
10 crescentus strains (see: MacRae and Smit (1991) [supra],
and Walker, S.G. et al (1992) [supra]). Alternatively, one
may use recombinant fusion S-layer genes produced with
C. crescentus S-layer gene and express them in alternate
Caulobacter hosts.

15 Freshwater Caulobacter producing S-layers may be readily
detected by negative stain transmission electron microscopy
techniques. Caulobacter may be isolated using the methods
outlined by MacRae and Smit (1991) [supra], which take
20 advantage of the fact that Caulobacter can tolerate periods
of starvation while other soil and water bacteria may not
and that they all produce a distinctive stalk structure,
visible by light microscopy (using either phase contrast or
standard dye staining methods). Once Caulobacter strains
25 are isolated in a typical procedure, colonies are suspended
in 2% ammonium molybdate negative stain and applied to
plastic-filmed, carbon-stabilized 300 or 400 mesh copper or
nickel grids and examined in a transmission electron
microscope at 60 kilovolt accelerating voltage (see: Smit,
30 J. (1986) "Protein Surface Layers of Bacteria", in Outer
Membranes as Model Systems, (M. Inouge, ed. J.Wiley & Sons,
at p. 343-376). S-layers are seen as two-dimensional
geometric patterns most readily on those cells in a colony
that have lysed and released their internal contents.

35

-16-

Recombinant Prot in Purification

Secreted proteins are separated and "shed" into the culture media as a macroscopic precipitate (the aggregate described above). The shedding phenomenon is a consequence of the absence of the N-terminal region of the S-layer protein in the expressed recombinant protein, or the loss of the lipopolysaccharide species used for S-layer attachment by the Caulobacter (see: Walker, S.G. et al (1994) [*supra*]).

The loose gel like clumps of aggregate may be readily separated from a soluble cleaved target protein by any suitable techniques such as filtration or centrifugation. If the target protein is insoluble once cleaved, it may then be convenient to solubilize one or both of the proteins (for example in 8M urea or 6M guanidine HCL) and separate by chromatography. In this way, only 2 species of protein need to be separated.

Cleavage of Fusion Proteins

Conditions for cleavage at aspartate - proline sites are described in Current Protocols in Molecular Biology (supp. 28; chapter 16.4) John Wiley & Sons Inc. 1994, and in Landon, M. "Cleavage at Aspartyl - Prolyl Bonds" in Methods in Enzymology (1977) 47: 145-149. These references suggest that significant variability of cleavage conditions exist for different proteins and that cleavage might occur in some instances without first denaturing or solubilizing the protein. However, in practice the latter circumstances are rare and proteins to be subjected to acid cleavage at Asp-Pro dipeptides are usually solubilized to a state where there is no visible turbidity. The solubilized protein will normally not pellet when centrifuged at 100,000 x g for 1 hour. It is now shown that mild-acid conditions may be used for cleavage of aspartate-proline sites in

-17-

Caulobacter S-layer fusion proteins without placing the protein in a solubilized state as described above.

5 In the method of this invention, conditions are adjusted to avoid destruction of the target protein or solubilization of the aggregate containing the S-layer secretion signal. Excess acid or too high a temperature may increase the occurrence over time of random cleavages along the length of the fusion protein, which is to be avoided since such
10 random cleavages may lead to undersized fragmentation of the fusion protein or solubilization of the aggregated S-layer portion.

15 Good yields of target protein with minimum random breaks in the fusion protein may generally be achieved by using from 5-20 mM HCL (or its equivalent while employing another acid). The respective pH of these conditions (unbuffered acid solution) is from about 2.3 to about 1.69. Time and temperature is preferably adjusted to achieve the desired
20 cleavage while minimizing random breaks. For example, temperature may range from room temperature to about 50° C. Time of treatment may range from about 12-72 hours. Time or temperature outside of these ranges is permissible depending upon the strength of the acid and the accepted
25 yield. Generally, lower yields are obtained with less acid strength, less time or lower temperatures.

In the following examples, efficiency of cleavage in the order of 40-80% is achieved using conditions similar to the
30 following alternatives:

- 5 mM HCL at 50° C. for 48-72 hours
- 20 mM HCL at 30° C. for 48-72 hours.

35 Conditions in excess of the aforementioned values may be employed in some cases with the possibility of random breaks increasing particularly with acid strength or

-18-

temperature. In the following examples, significant random cleavage occurred with 50 mM HCL at 50° C. after 48 hours.

Any acid may be employed in this invention which is normally used in solutions to which proteins are exposed. Acids which have a deleterious effect on proteins under dilute conditions should be avoided. For example, HCL or an equivalent amount of H₂SO₄ may be used in this invention but oxidizing acids such as nitric acid may not be suitable.

Example 1. Cleavage of artificial silk protein sequences from a secretion signal containing a native aspartate-proline cleavage site.

An artificial protein sequence resembling spider silk was constructed by synthesis of partially overlapping and complementing oligomers of DNA, which were then completed to a full duplex DNA with TaqI polymerase extension, to create a sequence that coded for 97 amino acids. The resulting DNA sequence and corresponding amino acid sequence is shown in Appendix 2.

The DNA sequence shown in Appendix 2 was cloned into a gene carrier sequence residing in a pUC8 plasmid cloning vector. The gene segment carrier had BamHI restriction sites at each end and an internal BgIII site. This combination of restrictions sites allowed the production of multimers of the above sequence, relying on the fact that BamHI sticky ends will ligate into BgIII sticky end, with the loss of both restriction sites. Thus one copy of the silk-like sequence within the gene segment carrier can be put inside a second copy of the same to produce a dimer. Using this principle, an 8X repeat was produced, fused to DNA encoding the S-layer secretion signal corresponding to the C-terminal portion of the C. crescentus S-layer protein from about amino acid 690 onwards (see: Appendix 1). This

-19-

fusion protein gene was introduced into strain CB2A on a broad host range plasmid vector. The 8x multimer appeared to be unstable, resulting in recombination events that reduced the 8X multimer to a 3x size. The 3 fold repeat of the above 97 amino acid sequence, fused to the S-layer secretion signal was secreted. Protein was collected and subjected to treatment with 5mM HCL for 2 days at 50° C. The result was the liberation of about 80% of soluble silk-like polymer which was readily separated by filtration from the S-layer protein which remained completely aggregated under these conditions. Cleavage occurred at native aspartate-proline dimer in the Caulobacter S-layer signal region (see: Appendix 1, amino acids numbered 692-693).

Example 2. Cleavage of the salmonid virus Infectious Pancreatic Necrosis Virus (IPNV) surface glycoprotein candidate vaccine sequence from an S-layer secretion signal containing a native aspartate-proline site.

The surface glycoprotein of the IPNV strain is a vaccine candidate. For this example and Example 4, the sequence of the first 257 amino acids of the mature protein and the corresponding DNA sequence is shown in Appendix 3.

DNA encoding a segment of the major surface glycoprotein gene of IPNV specifying amino acids 145-257 of the protein was fused to DNA sequence specifying two putative T-cell activating epitopes: MVF (SEQ ID No:1; LSEIKGVIVHRLEGV, derived from Measles Virus protein F) and P2 (SEQ ID No:2; QYIKANSKFIGITEL, derived from tetanus toxoid protein). The T-cell epitopes were positioned on the C-terminal end of the IPNV sequence. This chimeric protein was in turn fused in frame with the C-crescentus S-layer gene at about amino acid 690 position of the gene and introduced into Caulobacter on a broad host range plasmid vector. The resulting secreted protein was collected and treated with

-20-

5 mM HCL for 2 days at 50° C. Cleavage occurred at the native aspartate-proline dimer described in Example 1. The result was the liberation of about 75% of soluble vaccine candidate chimeric protein from the S-layer secretion signal which remained aggregated.

Example 3. Cleavage of segments of an E. coli type I pilus tip subunit from an S-layer secretion signal containing a native aspartate-proline cleavage site.

10

The FimH gene product is the tip pilus subunit of the E. coli strains involved with urinary tract infections. Two segments, T3, specifying the first 145 amino acids of the mature peptide and T7, specifying the entire 258 amino acids of the mature peptide were fused to the S-layer secretion signal at about amino acid 690 of the S-layer sequence. The T3 and T7 sequences are shown in Appendix 4.

15

The fusion protein genes were introduced into strain CB2A on a broad host range plasmid vector. In both cases the resulting secreted protein was collected and treated with 5 mM HCL for 2 days at 50° C. In both cases, the result was the liberation of about 50% of soluble vaccine candidate chimeric protein from the S-layer secretion signal which remained aggregated. Cleavage occurred at the native aspartate-proline dimer described in Example 1.

20

25

Example 4. Cleavage of the salmonid virus IPNV surface glycoprotein candidate vaccine sequence from an S-layer secretion signal containing an introduced aspartate-proline cleavage site.

30

A segment of the major surface glycoprotein gene of IPNV specifying amino acids 1-257 of the protein shown in Appendix 4 was fused to a DNA sequence specifying a peptide containing an aspartate-proline dipeptide (SEQ ID No: 3; SPLGPAGDPEAS) such that the aspartate-proline dipeptide was

35

-21-

positioned very near the C-terminus of the chimeric protein. This chimeric protein was in turn fused in frame with the C. crescentus S-layer gene at about amino acid 784 position of the gene and introduced in strain CB2A on a
5 broad host range plasmid vector. The resulting secreted protein was collected and treated with 5 mM HCL for 2 days at 50° C. Cleavage occurred at the introduced aspartate-proline dipeptide. The result was the liberation
10 of about 40% of insoluble vaccine candidate chimeric protein from the S-layer secretion signal which remained aggregated.

Lengthy DNA and amino acid sequences referred to in this specification are set out in the following appendixes.
15 Appendix 1 sets out the complete nucleotide sequence of the C. crescentus S-layer gene (SEQ ID No: 4) with the upstream sequence including the -35 and -10 sites of the promoter region and the Shine Dalgarno sequence. The start codon is at nucleotide 101 and the coding sequence run to and
20 includes nucleotide 3179. The amino acid sequence of the C. crescentus S-layer protein (SEQ ID No: 5) included in Appendix 1 is predicted from the DNA sequence. Appendix 2 sets out the artificial spider silk DNA sequence (SEQ ID No:6) used in Example 1 and the corresponding amino acid
25 sequence (SEQ ID No. 7). Appendix 3 sets out the DNA sequence (SEQ ID No: 8) and corresponding amino acid sequence (SEQ ID No: 9) of the first 257 amino acids of IPNV as described in Examples 2 and 4. Appendix 4 sets out the T3 protein sequence (SEQ ID No: 10) and the T7 protein
30 sequence (SEQ ID No: 11) as described in Example 3.

All publications, patents and patent applications referred to herein are hereby incorporated by reference. While this invention has been described according to particular
35 embodiments and by reference to certain examples, it will be apparent to those of skill in the art that variations

-22-

and modifications of the invention as described herein fall within the spirit and scope of the attached claims.

Appendix 1

GCTATTGTCG	ACGTATGACG	TTTGCTCTAT	AGCCATCGCT	GCTCCCATGC	GCGCCACTCG	60
GTGCGAGGGG	GTGTGGGATT	TTTTTTGGGA	GACAATCCTC	ATGGCCTATA	CGACGGCCCA	120
GTGGGTGACT	GCGTACACCA	ACGCCAACCT	CGGCAAGGCG	CCTGACGCCG	CCACCACGCT	180
GACGCTCGAC	GCGTACGCGA	CTCAAACCCA	GACGGGCGGC	CTCTCGGACG	CCGCTGCGCT	240
GACCAACACC	CTGAAGCTGG	TCAACAGCAC	GACGGCTGTT	GCCATCCAGA	CCTACCAGTT	300
CTTCACCGGC	GTTGCCCCGT	CGGCCGCTGG	TCTGGACTTC	CTGGTCGACT	CGACCACCAA	360
CACCAACGAC	CTGAACGACG	CGTACTACTC	GAAGTTTCGT	CAGGAAAACC	GCTTCATCAA	420
CTTCTCGATC	AACCTGGCCA	CGGGCGCCGG	CGCCGGCGCG	ACGGCTTTTCG	CCGCCGCCTA	480
CACGGGCGTT	TCGTACGCCC	AGACGGTCGC	CACCGCCTAT	GACAAGATCA	TCGGCAACGC	540
CGTCGCGACC	GCCGCTGGCG	TCGACGTCCG	GGCCGCCGTG	GCTTTCCTGA	GCCGCCAGGC	600
CAACATCGAC	TACCTGACCG	CCTTCGTGCG	CGCCAACACG	CCGTTACCGG	CCGCTGCCGA	660
CATCGATCTG	GCCGTCAAGG	CCGCCCTGAT	CGGCACCATC	CTGAACGCCG	CCACGGTGTC	720
GGGCATCGGT	GGTTACGCGA	CCGCCACGGC	CGCGATGATC	AACGACCTGT	CGGACGGCGC	780
CCTGTGCGAC	GACAACGCGG	CTGGCGTGAA	CCTGTTTACC	GCCTATCCGT	CGTCGGGCGT	840
GTGCGGTTTC	ACCCTCTCGC	TGACCACCGG	CACCGACACC	CTGACGGGCA	CCGCCAACAA	900
CGACACGTTT	GTTGCGGGTG	AAGTCGCCGG	CGTCGCGACC	CTGACCGTTG	GCGACACCCT	960
GAGCGGCGGT	GCTGGCACCG	ACGTCCTGAA	CTGGGTGCAA	GCTGCTGCGG	TTACGGCTCT	1020
GCCGACCGGC	GTGACGATCT	CGGGCATCGA	AACGATGAAC	GTGACGTCGG	GCGCTGCGAT	1080
CACCCTGAAC	ACGTCTTCGG	GCGTGACGGG	TCTGACCGCC	CTGAACACCA	ACACCAGCGG	1140
CGCGGCTCAA	ACCGTCACCG	CCGGCGCTGG	CCAGAACCTG	ACCGCCACGA	CCGCCGCTCA	1200
AGCCGCGAAC	AACGTCGCCG	TCGACGGGCG	CGCCAACGTC	ACCGTCGCCT	CGACGGGCGT	1260
GACCTCGGGC	ACGACCACGG	TCGGCGCCAA	CTCGGCCGCT	TCGGGCACCG	TGTCGGTGAG	1320
CGTCGCGAAC	TCGAGCACGA	CCACCACGGG	CGCTATCGCC	GTGACCGGTG	GTACGGCCGT	1380
GACCGTGGCT	CAAACGGCCG	GCAACGCCGT	GAACACCACG	TTGACGCAAG	CCGACGTGAC	1440
CGTGACCGGT	AACTCCAGCA	CCACGGCCGT	GACGGTCACC	CAAACGCCCG	CCGCCACCGC	1500
CGGCGCTACG	GTCGCCGGTC	GCGTCAACGG	CGCTGTGACG	ATCACCAGCT	CTGCCGCCGC	1560
CTCGGCCACG	ACCGCCGGCA	AGATCGCCAC	GGTCACCCTG	GGCAGCTTCG	GCGCCGCCAC	1620
GATCGACTCG	AGCGCTCTGA	CGACCGTCAA	CCTGTGCGGC	ACGGGCACCT	CGCTCGGCAT	1680

Appendix 1
(cont'd)

CGGCCGCGGC	GCTCTGACCG	CCACGCCGAC	CGCCAACACC	CTGACCCTGA	ACGTCAATGG	1740
TCTGACGACG	ACCGGCGCGA	TCACGGACTC	GGAAGCGGCT	GCTGACGATG	GTTTCACCAC	1800
CATCAACATC	GCTGGTTCGA	CCGCCTCTTC	GACGATCGCC	AGCCTGGTGG	CCGCCGACGC	1860
GACGACCCTG	AACATCTCGG	GCGACGCTCG	CGTCACGATC	ACCTCGCACA	CCGCTGCCGC	1920
CCTGACGGGC	ATCACGGTGA	CCAACAGCGT	TGGTGCGACC	CTCGGCGCCG	AACTGGCGAC	1980
CGGTCTGGTC	TTCACGGGCG	GCGCTGGCCG	TGACTCGATC	CTGCTGGGCG	CCACGACCAA	2040
GGCGATCGTC	ATGGGCGCCG	GCGACGACAC	CGTCACCGTC	AGCTCGGCGA	CCCTGGGCGC	2100
TGGTGGTTTC	GTCAACGGCG	GCGACGGCAC	CGACGTTCTG	GTGGCCAACG	TCAACGGTTC	2160
GTCGTTCAAG	GCTGACCCGG	CCTTCGGCCG	CTTCGAAACC	CTCCGCGTCG	CTGGCGCGGC	2220
GGCTCAAGGC	TCGCACAACG	CCAACGGCTT	CACGGCTCTG	CAACTGGGCG	CGACGGCGGG	2280
TGCGACGACC	TTCACCAACG	TTGCGGTGAA	TGTCGGCCTG	ACCGTTCTGG	CGGCTCCGAC	2340
CGGTACGACG	ACCGTGACCC	TGGCCAACGC	CACGGGCACC	TCGGACGTGT	TCAACCTGAC	2400
CCTGTCGTCC	TCGGCCGCTC	TGGCCGCTGG	TACGGTTGCG	CTGGCTGGCG	TCGAGACGGT	2460
GAACATCGCC	GCCACCGACA	CCAACACGAC	CGCTCACGTC	GACACGCTGA	CGCTGCAAGC	2520
CACCTCGGCC	AAGTCGATCG	TGGTGACGGG	CAACGCCGGT	CTGAACCTGA	CCAACACCGG	2580
CAACACGGCT	GTCAACAGCT	TCGACGCCAG	CGCCGTCACC	GGCACGGCTC	CGGCTGTGAC	2640
CTTCGTGTCT	GCCAACACCA	CGGTGGGTGA	AGTCGTCACG	ATCCGCGGCG	GCGCTGGCGC	2700
CGACTCGCTG	ACCGGTTCCG	CCACCGCCAA	TGACACCATC	ATCGGTGGCG	CTGGCGCTGA	2760
CACCCTGGTC	TACACCGGCG	GTACGGACAC	CTTCACGGGT	GGCACGGGCG	CGGATATCTT	2820
CGATATCAAC	GCTATCGGCA	CCTCGACCGC	TTCGTGACG	ATCACCGACG	CCGCTGTCCG	2880
CGACAAGCTC	GACCTCGTCG	GCATCTCGAC	GAACGGCGCT	ATCGCTGACG	GCGCCTTCGG	2940
CGCTGCGGTC	ACCCTGGGCG	CTGCTGCGAC	CCTGGCTCAG	TACCTGGACG	CTGCTGCTGC	3000
CGGCGACGGC	AGCGGCACCT	CGGTTGCCAA	GTGGTTCCAG	TTCGGCGGCG	ACACCTATGT	3060
CGTCGTTGAC	AGCTCGGCTG	GCGCGACCTT	CGTCAGCGGC	GCTGACGCGG	TGATCAAGCT	3120
GACCGGTCTG	GTCACGCTGA	CCACCTCGGC	CTTCGCCACC	GAAGTCCTGA	CGCTCGCCTA	3180
AGCGAACGTC	TGATCCTCGC	CTAGGCGAGG	ATCGCTAGAC	TAAGAGACCC	CGTCTTCCGA	3240
AAGGGAGGCG	GGGTCTTTCT	TATGGGCGCT	ACGCGCTGGC	CGGCCTTGCC	TAGTTCCGGT	3300

Appendix 1
(cont'd)

Met 1	Ala	Tyr	Thr	Thr 5	Ala	Gln	Leu	Val	Thr 10	Ala	Tyr	Thr	Asn 15	Ala	Asn
Leu	Gly	Lys	Ala 20	Pro	Asp	Ala	Ala	Thr 25	Thr	Leu	Thr	Leu	Asp 30	Ala	Tyr
Ala	Thr	Gln	Thr	Gln	Thr	Gly	Gly 40	Leu	Ser	Asp	Ala	Ala	Ala 45	Leu	Thr
Asn	Thr	Leu	Lys	Leu	Val	Asn 55	Ser	Thr	Thr	Ala	Val 60	Ala	Ile	Gln	Thr
Tyr 65	Gln	Phe	Phe	Thr	Gly 70	Val	Ala	Pro	Ser	Ala 75	Ala	Gly	Leu	Asp	Phe 80
Leu	Val	Asp	Ser	Thr 85	Thr	Asn	Thr	Asn	Asp 90	Leu	Asn	Asp	Ala	Tyr	Tyr 95
Ser	Lys	Phe	Ala 100	Gln	Glu	Asn	Arg	Phe 105	Ile	Asn	Phe	Ser	Ile 110	Asn	Leu
Ala	Thr	Gly	Ala	Gly	Ala	Gly	Ala 120	Thr	Ala	Phe	Ala	Ala 125	Ala	Tyr	Thr
Gly	Val 130	Ser	Tyr	Ala	Gln	Thr 135	Val	Ala	Thr	Ala	Tyr 140	Asp	Lys	Ile	Ile
Gly 145	Asn	Ala	Val	Ala	Thr 150	Ala	Ala	Gly	Val	Asp 155	Val	Ala	Ala	Ala	Val 160
Ala	Phe	Leu	Ser	Arg 165	Gln	Ala	Asn	Ile	Asp 170	Tyr	Leu	Thr	Ala	Phe 175	Val
Arg	Ala	Asn	Thr 180	Pro	Phe	Thr	Ala	Ala 185	Ala	Asp	Ile	Asp	Leu 190	Ala	Val
Lys	Ala	Ala 195	Leu	Ile	Gly	Thr	Ile 200	Leu	Asn	Ala	Ala	Thr 205	Val	Ser	Gly
Ile	Gly 210	Gly	Tyr	Ala	Thr	Ala 215	Thr	Ala	Ala	Met	Ile 220	Asn	Asp	Leu	Ser
Asp 225	Gly	Ala	Leu	Ser	Thr 230	Asp	Asn	Ala	Ala	Gly 235	Val	Asn	Leu	Phe	Thr 240
Ala	Tyr	Pro	Ser	Ser 245	Gly	Val	Ser	Gly	Ser 250	Thr	Leu	Ser	Leu	Thr 255	Thr
Gly	Thr	Asp	Thr 260	Leu	Thr	Gly	Thr	Ala 265	Asn	Asn	Asp	Thr	Phe 270	Val	Ala
Gly	Glu	Val 275	Ala	Gly	Ala	Ala	Thr 280	Leu	Thr	Val	Gly	Asp 285	Thr	Leu	Ser
Gly 290	Gly	Ala	Gly	Thr	Asp	Val 295	Leu	Asn	Trp	Val	Gln 300	Ala	Ala	Ala	Val
Thr 305	Ala	Leu	Pro	Thr	Gly 310	Val	Thr	Ile	Ser	Gly 315	Ile	Glu	Thr	Met	Asn 320
Val	Thr	Ser	Gly	Ala 325	Ala	Ile	Thr	Leu	Asn 330	Thr	Ser	Ser	Gly	Val 335	Thr
Gly	Leu	Thr	Ala 340	Leu	Asn	Thr	Asn	Thr 345	Ser	Gly	Ala	Ala	Gln 350	Thr	Val

Appendix 1
(cont'd)

Thr Ala Gly Ala Gly Gln Asn Leu Thr Ala Thr Thr Ala Ala Gln Ala
 355 360 365
 Ala Asn Asn Val Ala Val Asp Gly Arg Ala Asn Val Thr Val Ala Ser
 370 375 380
 Thr Gly Val Thr Ser Gly Thr Thr Thr Val Gly Ala Asn Ser Ala Ala
 385 390 395 400
 Ser Gly Thr Val Ser Val Ser Val Ala Asn Ser Ser Thr Thr Thr Thr
 405 410 415
 Gly Ala Ile Ala Val Thr Gly Gly Thr Ala Val Thr Val Ala Gln Thr
 420 425 430
 Ala Gly Asn Ala Val Asn Thr Thr Leu Thr Gln Ala Asp Val Thr Val
 435 440 445
 Thr Gly Asn Ser Ser Thr Thr Ala Val Thr Val Thr Gln Thr Ala Ala
 450 455 460
 Ala Thr Ala Gly Ala Thr Val Ala Gly Arg Val Asn Gly Ala Val Thr
 465 470 475 480
 Ile Thr Asp Ser Ala Ala Ala Ser Ala Thr Thr Ala Gly Lys Ile Ala
 485 490 495
 Thr Val Thr Leu Gly Ser Phe Gly Ala Ala Thr Ile Asp Ser Ser Ala
 500 505 510
 Leu Thr Thr Val Asn Leu Ser Gly Thr Gly Thr Ser Leu Gly Ile Gly
 515 520 525
 Arg Gly Ala Leu Thr Ala Thr Pro Thr Ala Asn Thr Leu Thr Leu Asn
 530 535 540
 Val Asn Gly Leu Thr Thr Thr Gly Ala Ile Thr Asp Ser Glu Ala Ala
 545 550 555 560
 Ala Asp Asp Gly Phe Thr Thr Ile Asn Ile Ala Gly Ser Thr Ala Ser
 565 570 575
 Ser Thr Ile Ala Ser Leu Val Ala Ala Asp Ala Thr Thr Leu Asn Ile
 580 585 590
 Ser Gly Asp Ala Arg Val Thr Ile Thr Ser His Thr Ala Ala Ala Leu
 595 600 605
 Thr Gly Ile Thr Val Thr Asn Ser Val Gly Ala Thr Leu Gly Ala Glu
 610 615 620
 Leu Ala Thr Gly Leu Val Phe Thr Gly Gly Ala Gly Arg Asp Ser Ile
 625 630 635 640
 Leu Leu Gly Ala Thr Thr Lys Ala Ile Val Met Gly Ala Gly Asp Asp
 645 650 655
 Thr Val Thr Val Ser Ser Ala Thr Leu Gly Ala Gly Gly Ser Val Asn
 660 665 670
 Gly Gly Asp Gly Thr Asp Val Leu Val Ala Asn Val Asn Gly Ser Ser
 675 680 685
 Phe S r Ala Asp Pro Ala Phe Gly Gly Phe Glu Thr Leu Arg Val Ala
 690 695 700

Appendix 1
(cont'd)

Gly Ala Ala Ala Gln Gly Ser His Asn Ala Asn Gly Phe Thr Ala Leu
 705 710 715 720
 Gln Leu Gly Ala Thr Ala Gly Ala Thr Thr Phe Thr Asn Val Ala Val
 725 730 735
 Asn Val Gly Leu Thr Val Leu Ala Ala Pro Thr Gly Thr Thr Thr Val
 740 745 750
 Thr Leu Ala Asn Ala Thr Gly Thr Ser Asp Val Phe Asn Leu Thr Leu
 755 760 765
 Ser Ser Ser Ala Ala Leu Ala Ala Gly Thr Val Ala Leu Ala Gly Val
 770 775 780
 Glu Thr Val Asn Ile Ala Ala Thr Asp Thr Asn Thr Thr Ala His Val
 785 790 795 800
 Asp Thr Leu Thr Leu Gln Ala Thr Ser Ala Lys Ser Ile Val Val Thr
 805 810 815
 Gly Asn Ala Gly Leu Asn Leu Thr Asn Thr Gly Asn Thr Ala Val Thr
 820 825 830
 Ser Phe Asp Ala Ser Ala Val Thr Gly Thr Ala Pro Ala Val Thr Phe
 835 840 845
 Val Ser Ala Asn Thr Thr Val Gly Glu Val Val Thr Ile Arg Gly Gly
 850 855 860
 Ala Gly Ala Asp Ser Leu Thr Gly Ser Ala Thr Ala Asn Asp Thr Ile
 865 870 875 880
 Ile Gly Gly Ala Gly Ala Asp Thr Leu Val Tyr Thr Gly Gly Thr Asp
 885 890 895
 Thr Phe Thr Gly Gly Thr Gly Ala Asp Ile Phe Asp Ile Asn Ala Ile
 900 905 910
 Gly Thr Ser Thr Ala Phe Val Thr Ile Thr Asp Ala Ala Val Gly Asp
 915 920 925
 Lys Leu Asp Leu Val Gly Ile Ser Thr Asn Gly Ala Ile Ala Asp Gly
 930 935 940
 Ala Phe Gly Ala Ala Val Thr Leu Gly Ala Ala Ala Thr Leu Ala Gln
 945 950 955 960
 Tyr Leu Asp Ala Ala Ala Ala Gly Asp Gly Ser Gly Thr Ser Val Ala
 965 970 975
 Lys Trp Phe Gln Phe Gly Gly Asp Thr Tyr Val Val Val Asp Ser Ser
 980 985 990
 Ala Gly Ala Thr Phe Val Ser Gly Ala Asp Ala Val Ile Lys Leu Thr
 995 1000 1005
 Gly Leu Val Thr Leu Thr Thr Ser Ala Phe Ala Thr Glu Val Leu Thr
 1010 1015 1020
 Leu Ala
 1025

Appendix 2

GAA TTC AGA TCT CAG GGC GCG GGG CAG GGT GGC TAT GGT GGG CTC GGC
TCG CAA GGC
GCT
E F R S Q G A G Q G G Y G G L G S Q G A
GGC CTG GGT GGC CAG GGC GCT GGC GCG GCC GCG GCC GCT GCG GCC GGT
GGC
G R G G Q G A G A A A A A A G G
GCT GGC CAG GGC GGG CTG GGC TCG CAG GGC GCC GGC CAA GGC GCT GGC
GCC GCG GCC
GCT
A G Q G G L G S Q G A G Q G A G A A A A
GCG GCC GGT GGC GCC GGC CAG GGT GGC TAC GGC GGC CTG GGC AGC CAG
GGC GCC GGT
CGC
A A G G A G Q G G Y G G L G S Q G A G R
GGC GGT CAG GGC GCC GGT GCC GCG GCC GCT GCG GCC GGT GGC GCT GGG
CAA GGC GGC TAC
G G Q G A G A A A A A A G G A G Q G G Y
GGC GGT CTG GGA TCC
G G L G S

Page 29
Appendix 3

1/1

atg aac aca aac aag gca acc gca act tac ttg aaa tcc att atg ctt cca gag act
gga
Met asn thr asn lys ala thr ala thr tyr leu lys ser ile met leu pro glu thr
gly
61/21

cca gca agc atc ccg gac gac ata acg gag aga cac atc tta aaa caa gag acc tcg
tca
pro ala ser ile pro asp asp ile thr glu arg his ile leu lys gln glu thr ser
ser
121/41

tac aac tta gag gtc tcc gaa tca gga agt ggc att ctt gtt tgt ttc cct ggg gca
cca
tyr asn leu glu val ser glu ser gly ser gly ile leu val cys phe pro gly ala
pro
181/61

ggc tca cgg atc ggt gca cac tac aga tgg aat grg aac cag acg ggg ctg gag ttc
gac
gly ser arg ile gly ala his tyr arg trp asn ala asn gln thr gly leu glu phe
asp
241/81

cag tgg ctg gag acg tcg cag gac ctg aag aaa gcc ttc aac tac ggg agg ctg atc
tca
gln trp leu glu thr ser gln asp leu lys lys ala phe asn tyr gly arg leu ile
ser
301/101

agg aaa tac gac att caa agc tcc aca cta ccg gcc ggt ctc tat gct ctg aac ggg
acg
arg lys tyr asp ile gln ser ser thr leu pro ala gly leu tyr ala leu asn gly
thr
361/121

ctc aac gct gcc acc ttc gaa ggc agt ctg tct gag gtg gag agc ctg acc tac aat
agc
leu asn ala ala thr phe glu gly ser leu ser glu val glu ser leu thr tyr asn
ser
421/141

ctg atg tcc cta act acg aac ccc cag gac aaa gcc aac aac cag ctg gtg acc aaa
gga
leu met ser leu thr thr asn pro gln asp lys ala asn asn gln leu val thr lys
gly
481/161

gtc acc gtc ctg aat cta cca aca ggg ttc gac aaa cca tac gtc cgc cta gag gac
gag
val thr val leu asn leu pro thr gly phe asp lys pro tyr val arg leu glu asp
glu
541/181

aca ccc cag ggt ctc cag tca atg aac ggg gcc agg atg agg tgc aca gct gca att
gca
thr pro gln gly leu gln ser met asn gly ala arg met arg cys thr ala ala ile
ala
601/201

cca cgg agg tac gag atc gac ctc cca tcc caa agc cta ccc ccc gtt cct gcg aca
gga
pro arg arg tyr glu ile asp leu pro ser gln ser leu pro pro val pro ala thr
gly
661/221

acc ctc acc act ctc tac gag gga aac gcc gac atc gtc agc tcc aca aca gtg acg
gga
thr leu thr thr leu tyr glu gly asn ala asp ile val ser ser thr thr val thr
gly
721/241

gac ata aac ttc agt ctg gca gaa cga ccc gca aac gag acc agg ttc gac ttc cag
ctg
asp ile asn phe ser l u ala glu arg pro ala asn glu thr arg phe asp phe gln
leu

Appendix 4

The T3 protein sequence is:

FACKTANGTAIPIGGGSANVYVNLAPVWVNGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSA
SYPFPTTSETPRVVYNSRTDKPWPVALYLPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQ
CDVSA

The T7 protein sequence is:

FACKTANGTAIPIGGGSANVYVNLAPVWVNGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSA
SYPFPTTSETPRVVYNSRTDKPWPVALYLPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQ
CDVSARDVTVTLPDYRGSVPIPLTVYCAKSQNLGYLSGTHADAGNSIFTNTASFSPAQGVG
GAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

-31-

WHAT IS CLAIMED IS:

1. A method of cleaving a fusion protein consisting of a first component comprising all or part of a Caulobacter S-layer protein including a Caulobacter C-terminal secretion signal, and a second component comprising a heterologous polypeptide expressed and secreted from Caulobacter wherein the fusion protein comprises at least one aspartate-proline dipeptide, and wherein the method comprises the step of combining said fusion protein with an acid solution of a strength insufficient to solubilize the fusion protein for a time sufficient for cleavage of the fusion protein at said aspartate-proline dipeptide.
2. The method of claim 1 wherein the at least one dipeptide is situated between the first and second components or adjacent a junction between the first and second components.
3. The method of claim 1 wherein the acid solution has a pH of from about 1.5 to about 2.5.
4. The method of claim 1 wherein the acid solution has a pH of about 1.65 to about 2.35.
5. The method of any one of claims 1-4 wherein the method is carried out at a temperature in the range of about 30° C. to about 50° C.
6. The method of any one of claims 1-5 comprising the additional step of separating products cleaved from the fusion protein.
7. A method of preparing a DNA construct for expression of a fusion protein suitable for use in the method of claim 1, comprising joining an upstream DNA segment

Folder

-32-

5 comprising DNA encoding a heterologous protein of interest and a downstream DNA segment for a Caulobacter C-terminal secretion signal which does not encode an aspartate-proline dipeptide, wherein the upstream segment comprises a sequence encoding an aspartate-proline dipeptide at or near an end of said upstream segment to be joined to said downstream segment.

10 8. A method of preparing a fusion protein, comprising the steps of expressing a DNA construct prepared as described in claim 7 in Caulobacter and, recovering said fusion protein once secreted by the Caulobacter.